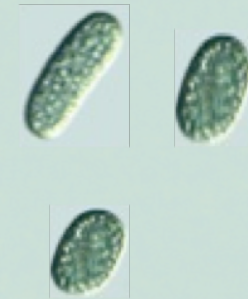
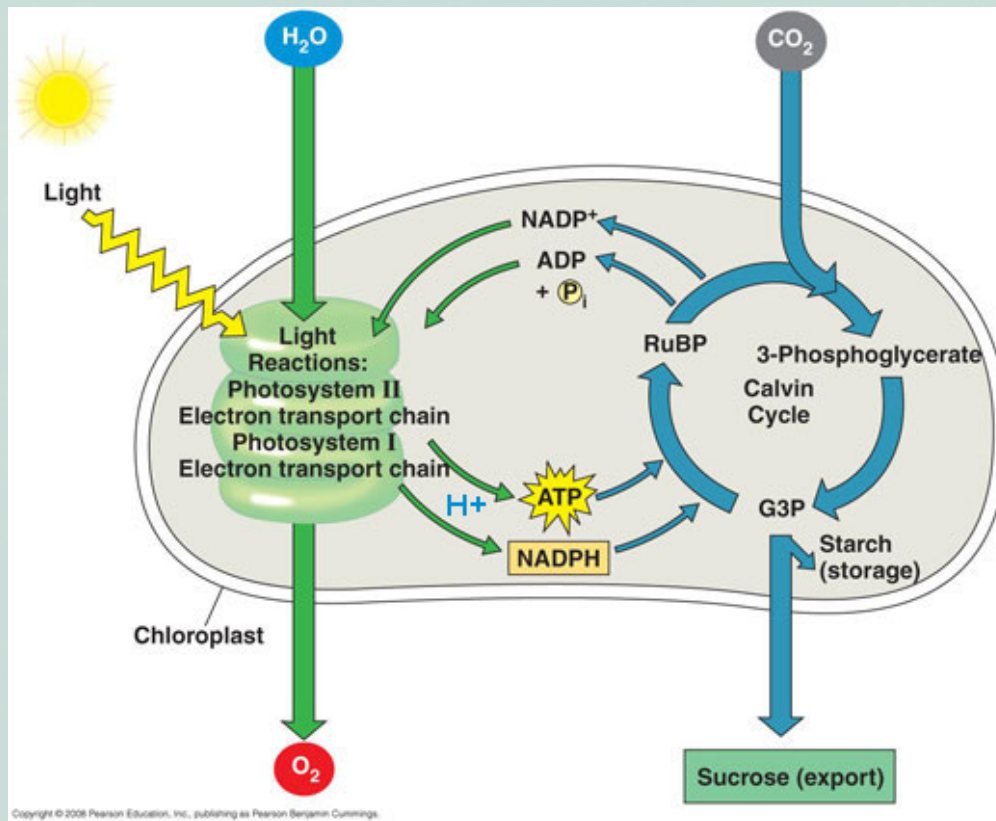


Fun with presence/absence data

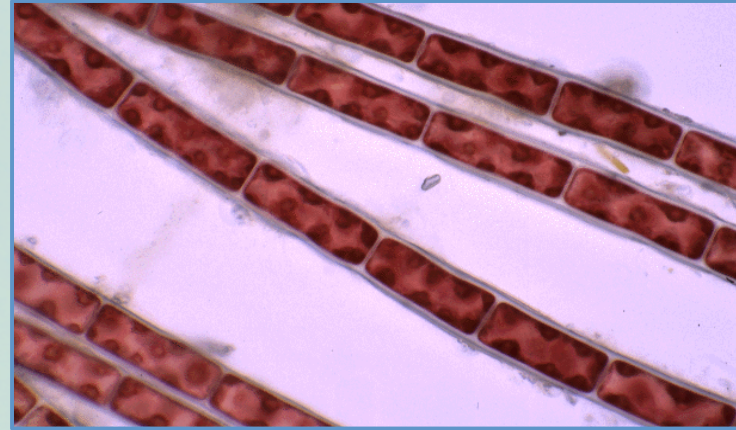
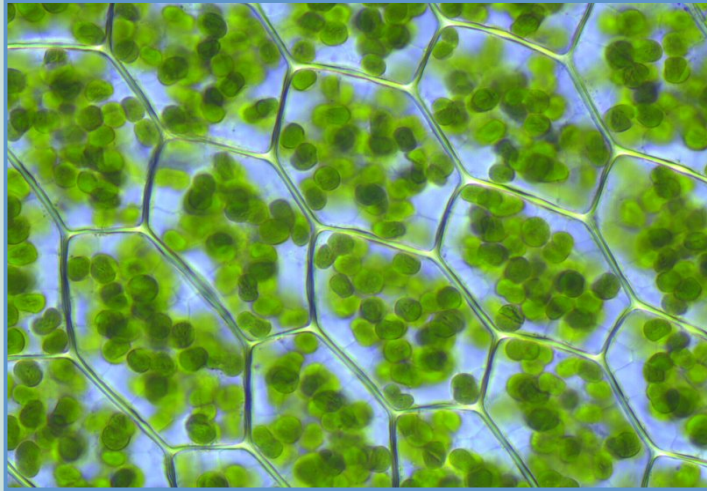
Barbara Schönfeld



Photosynthesis



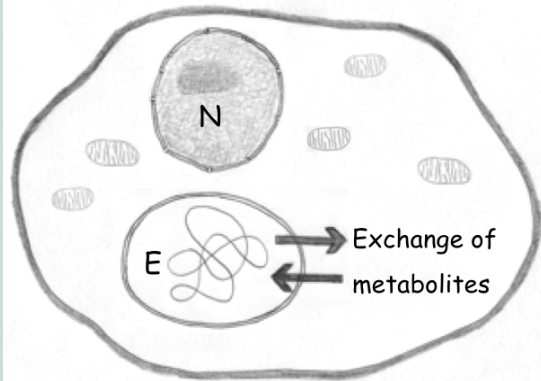
Plastids



Symbiogenesis

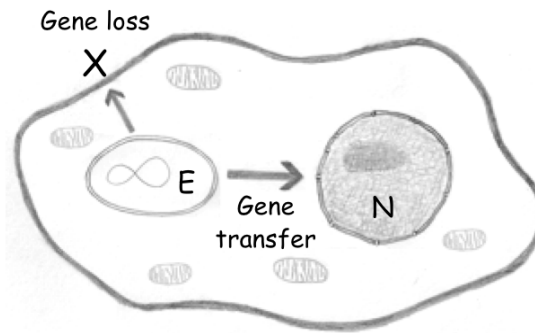
The stages of symbiogenesis:

E: Endosymbiont; N: Host nucleus; O: Organelle



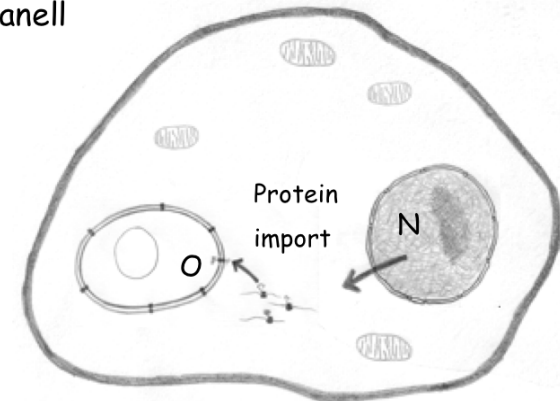
Facultative endosymbiosis

Symbiont and host have ability to live independently but benefit from stable exchange of metabolites



Obligate endosymbiosis

with concomitant genome reduction in the endosymbiont;



Eukaryote cell with organelle

a protein import system evolved to transfer nuclear expressed gene products into the former symbiont

Plastid encoded proteins

~ 2000 different proteins are necessary to run a plastid

Contemporary plastid genomes encode 20 – 200 proteins, the rest is encoded in the nucleus and imported

Things you'd think we'd know

What is the nature of the selection pressure to reduce the plastid genome?

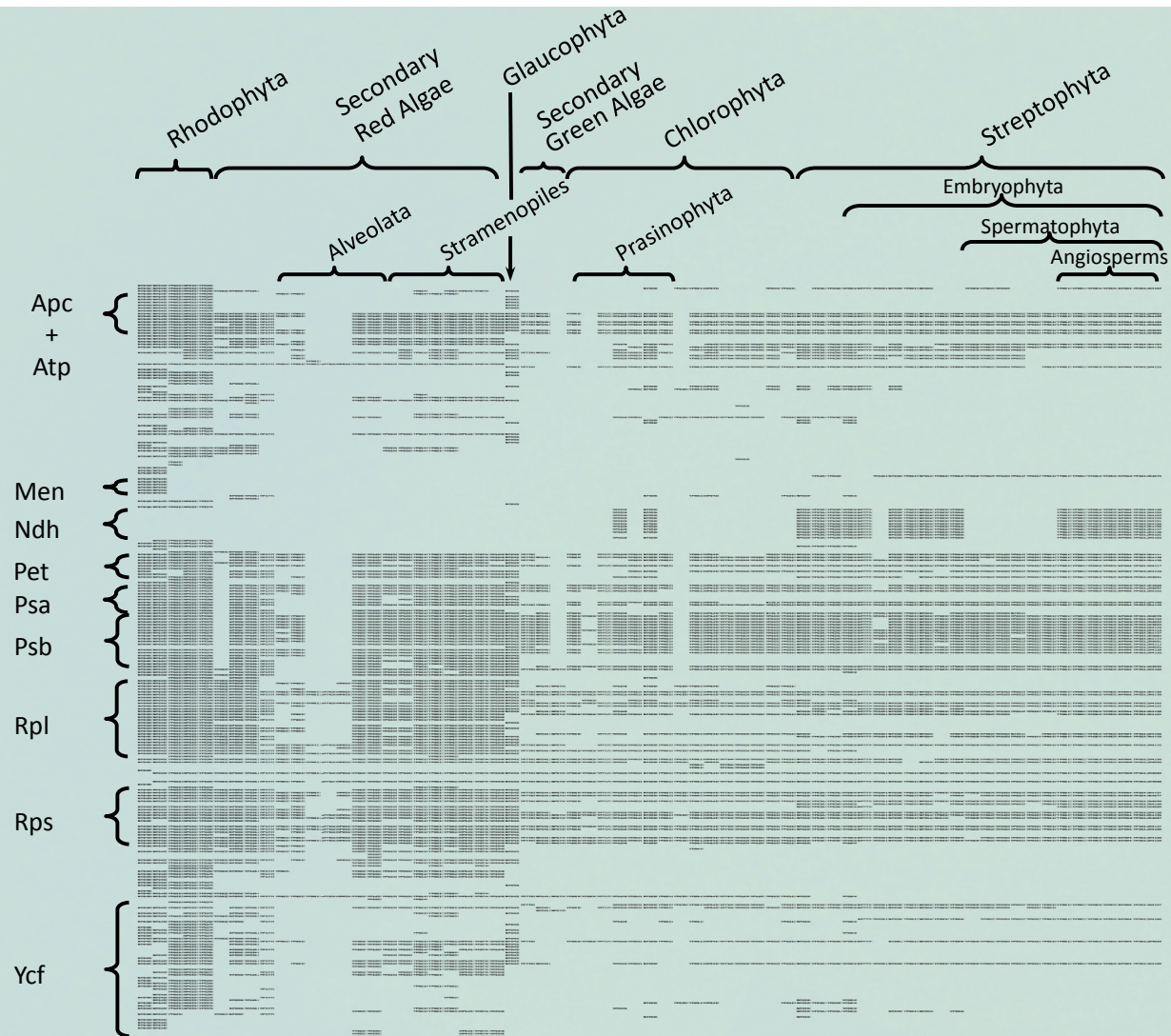
What keeps some genes in the plastid genome? Why are there plastid genomes left at all?

How prevalent are gene losses from the plastid genome?

Is the loss of specific genes primarily driven by chance or selection pressures?

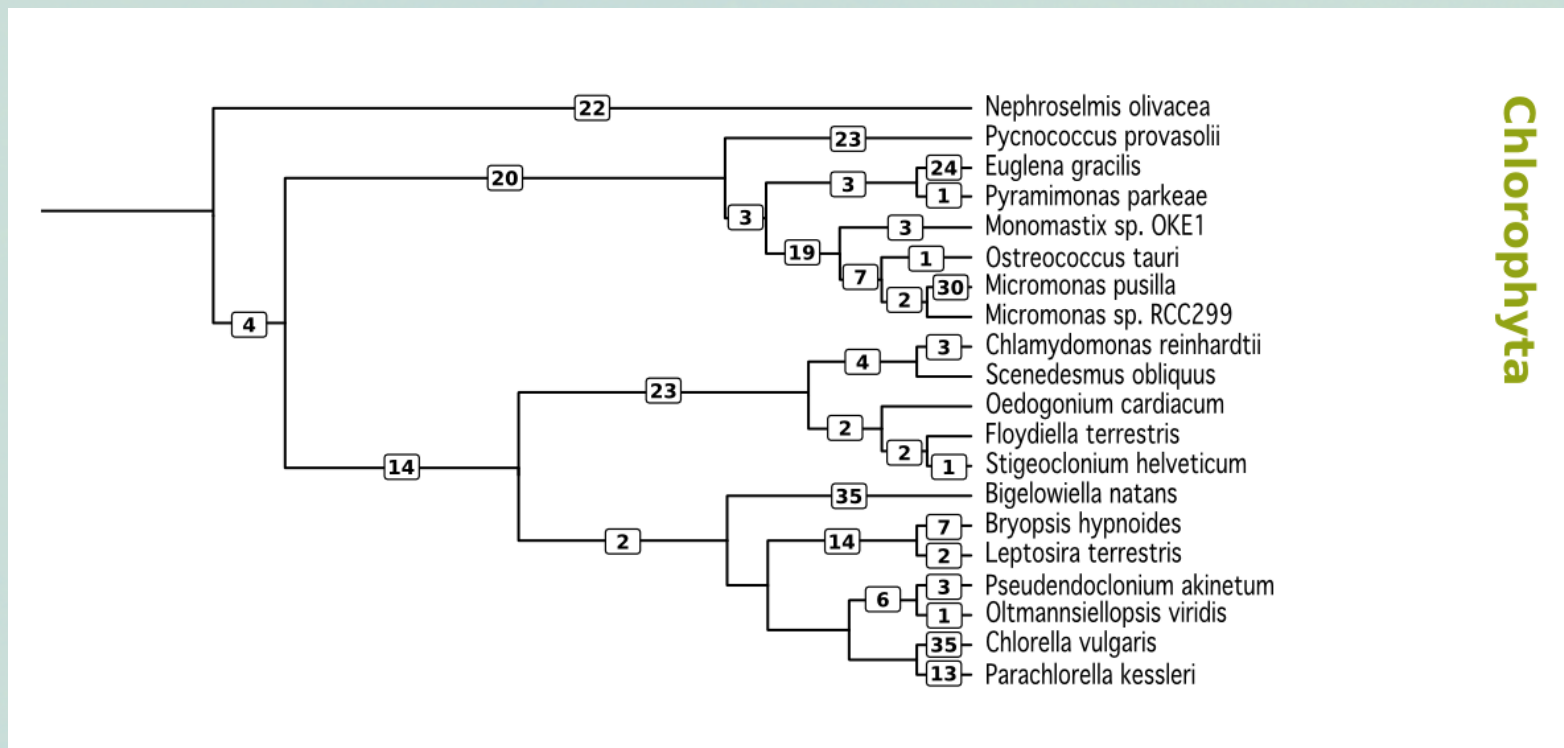
The data

196 Taxa
267 Genes



Dollo parsimony results

Gene losses happen more frequently than expected by many biologists.
There are lineage specific patterns and stark differences in frequency.



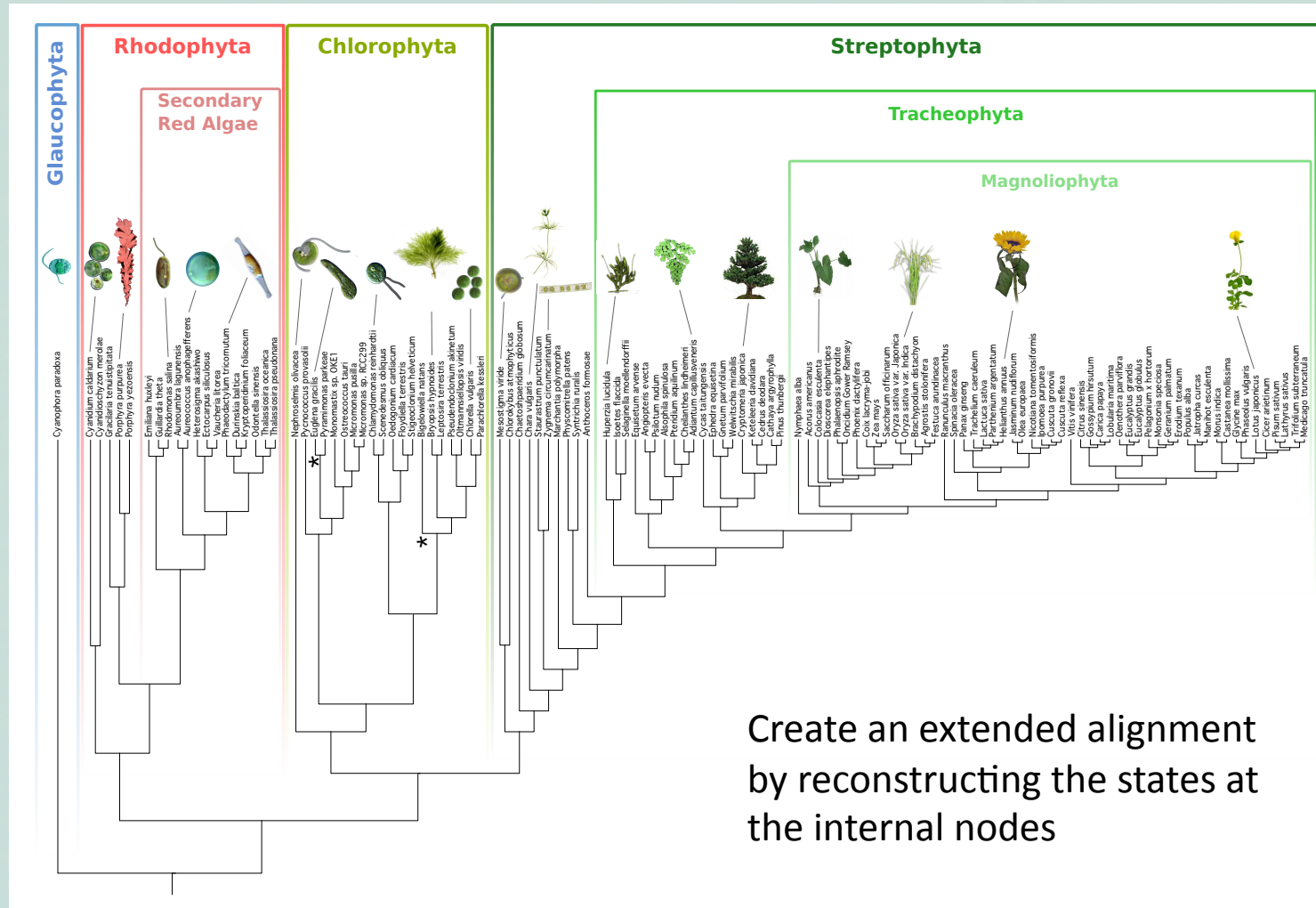
Was that it?

What frequency distribution should we expect to observe at random?

Does the loss of a specific gene change the probability of another gene being lost?

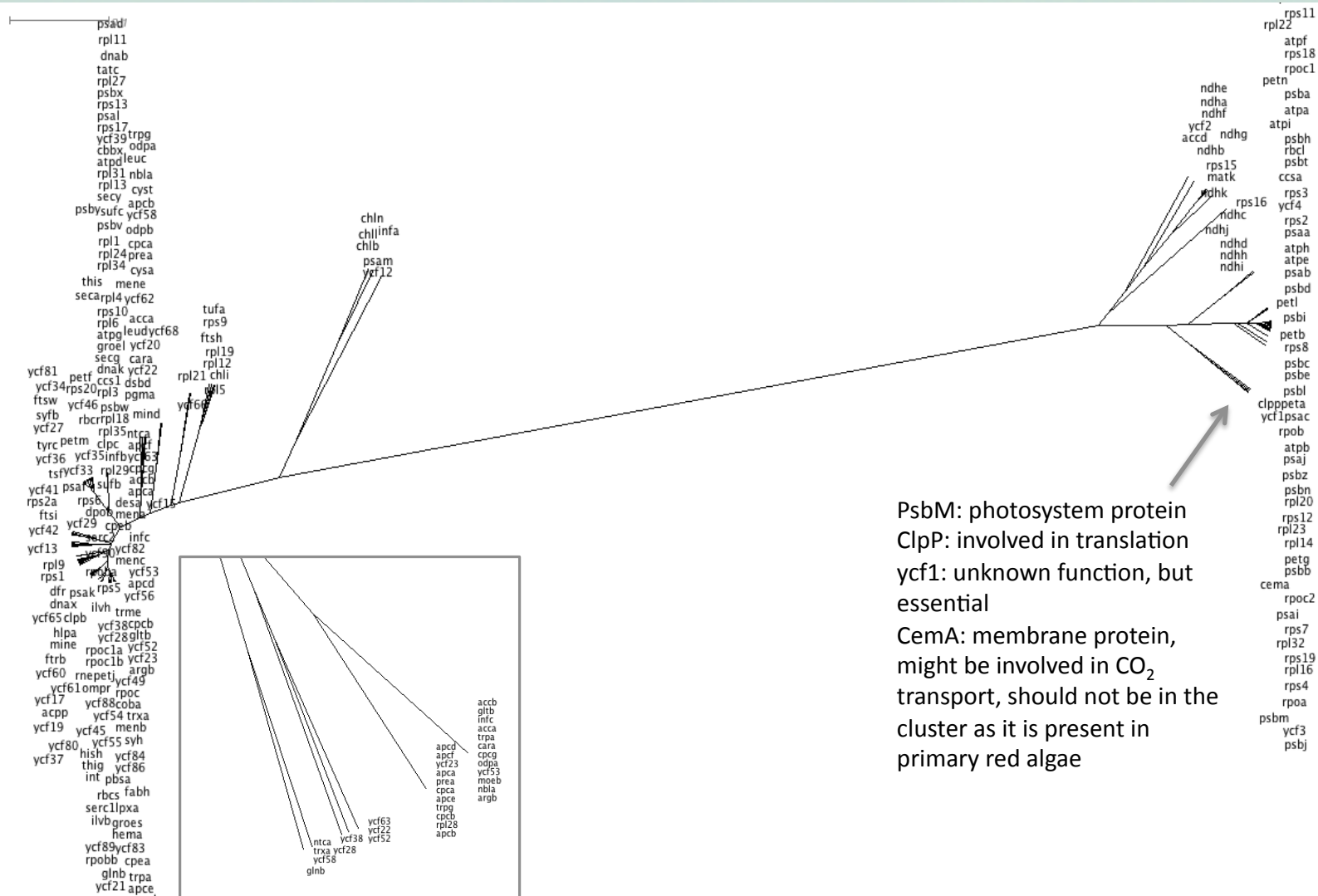
Do similarities in gene fate tell us anything about the underlying mechanisms?

Which genes share similar fates?



Create an extended alignment by reconstructing the states at the internal nodes

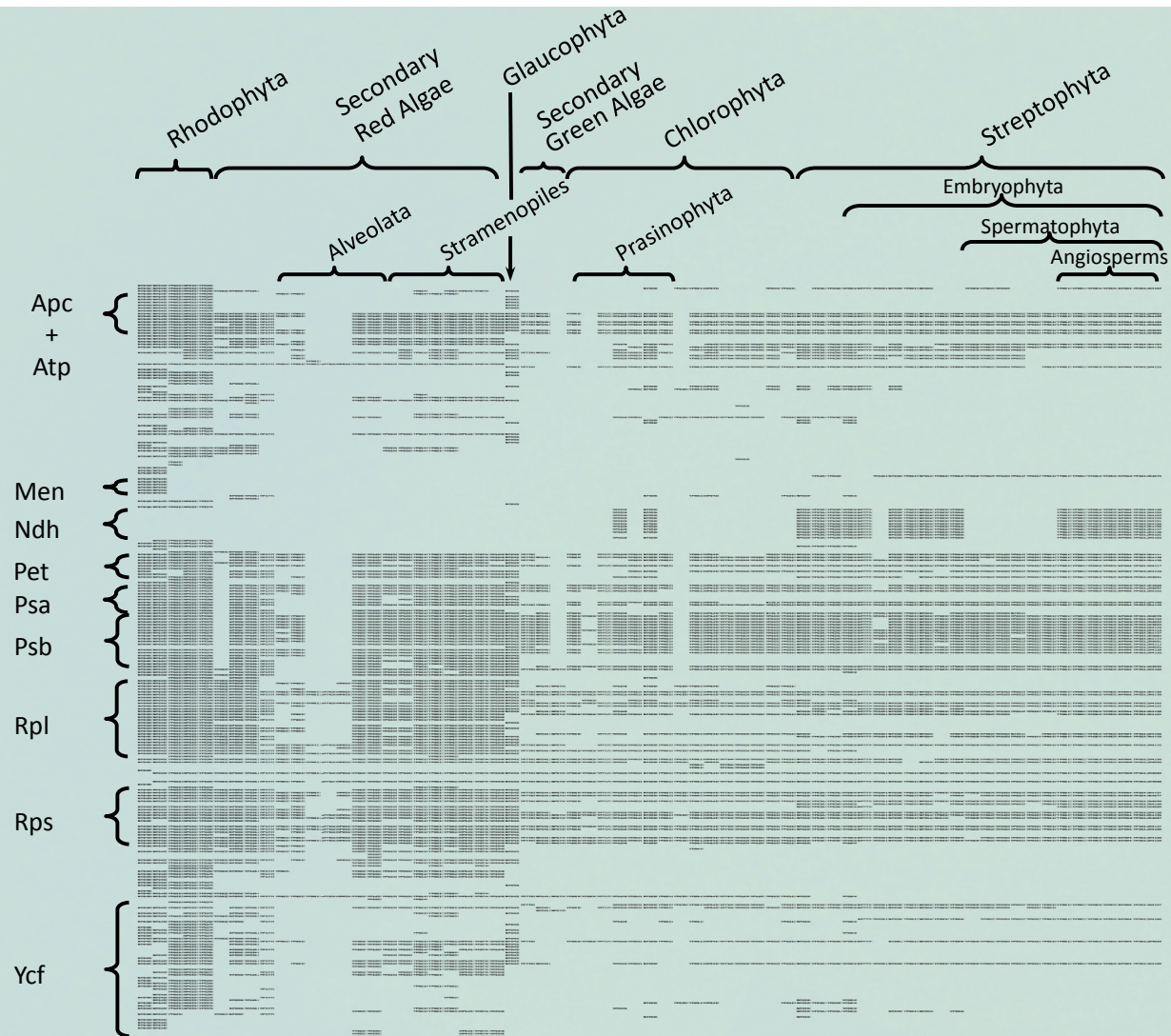
UPGMA of extended alignment



PsbM: photosystem protein
 ClpP: involved in translation
 ycf1: unknown function, but essential
 Cema: membrane protein, might be involved in CO₂ transport, should not be in the cluster as it is present in primary red algae

The data

196 Taxa
267 Genes



Acknowledgements

Thank you for listening!

I'd like to thank for their support:

Pete Lockhart

Bill Martin

Barbara Holland

Michael Woodhams

and

UTAS department of Math and

Physics for hosting me

